Relating Bisimulations with Attractors in Boolean Network Models

Daniel Figueiredo*

CIDMA - Center for Research and Development in Mathematics and Applications
Department of Mathematics, University of Aveiro
Campus Universitário de Santiago, 3810-193, Aveiro, Portugal
daniel.figueiredo@ua.pt

Abstract. When studying a biological regulatory network, it is usual to use boolean network models. In these models, boolean variables represent the behavior of each component of the biological system. Taking in account that the size of these state transition models grows exponentially along with the number of components considered, it becomes important to have tools to minimize such models. In this paper, we relate bisimulations, which are relations used in the study of automata (general state transition models) with attractors, which are an important feature of biological boolean models. Hence, we support the idea that bisimulations can be important tools in the study some main features of boolean network models. We also discuss the differences between using this approach and other well-known methodologies to study this kind of systems and we illustrate it with some examples.

Keywords: Biological regulatory networks; Bisimulation; Minimization of models.

1 Introduction

The term "biological regulatory network" refers to the regulation processes which occur within a cell. In this environment, there are several biological components which react with each other (for example, by chemical reactions). More generally, the occurrence of these regulation processes within a biological system is due to the fact that the presence of some components in the cell can either induce or inhibit the production of some other component(s). For instance, this behavior can be observed when some proteins interact with genes producing mRNA. In its turn, mRNA induce the production of other proteins and so on.

To study a biological regulatory network, we must take in account that state variables like the concentration of proteins, mRNA and other components vary

* This work was supported in part by the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), through CIDMA - Center for Research and Development in Mathematics and Applications, within project UID/MAT/04106/2013. The author also acknowledges the support of FCT via the Ph.D. scholarship PD/BD/114186/2016.

in a continuous form. Indeed, one of the most precise kind of models used in this field are those which describe the dynamics of a biological regulatory network by an ordinary differential equations system (see [5]) that only admits continuous state variables. Usually, these models use sigmoid functions to describe a positive/negative regulation of a component over another (*i.e.*, to describe that one component induces/inhibits the production of another). The sigmoid functions which are more often used to describe a positive regulation are the so called "Hill functions" and depend on parameters θ and n. In this way, it is not difficult to see that this kind of models admit non linear equations and, therefore, the resulting ordinary differential equations system is not trivially solved by analytic methods. Thus, other (more simple) kinds of models are often used in order to proceed with a preliminary study of the biological system. In this context, the boolean networks are really useful.

There are many variants of boolean network models, however, the basic idea of all them is to approach each state variable of the system by a boolean variable [9]. In this way, it is assigned either the value "1" or "0" to indicate that some component is present or absent, respectively. Then, for each component i we define a boolean variable x_i and consider a threshold θ_i. If the concentration of the component i is above θ_i we define x_i as "1" (present) and otherwise we define it as "0" (absent).

In a boolean network, a state is a vector $(x_1, ...,x_n)$ such that each x_i is the boolean variable associated to the component i. There also are some variants of these boolean approaches which admit more than one threshold associated to each component: θ_i^j. In this case, it is possible to obtain several levels of expressibility which are codified using several boolean variables x_i^j instead of only x_i. Still, there are other variants as asynchronous boolean networks (see [2]) which we will more carefully describe in this paper. A boolean network model which represents the dynamics of a biological system is a digraph in which each vertex is a vector composed of "0"s and "1"s (which relates to a possible configuration of the biological system); and each edge relates to a possible transition between states. We then have a state transition model (automata).

These biological models and their variants are widely used because they are simple and some features of the original system can be identified by studying these boolean models. One of the most studied features in biological regulatory networks is the existence of *steady states*. By steady state we refer to the values of the concentration of a cell's components where the system stabilizes. In particular, in models which use an ODE system, a steady state corresponds to those states in which the evolution of the system is null, *i.e.*, corresponds to set all differential equations to zero. When a steady state exists, it can be either stable (if little perturbations do not cause the system to evolve into a state far away from the initial steady state) or unstable (otherwise). Thus, the study of these characteristics is an important topic in the field of biological systems. Because of this, discrete models as boolean networks are often used because it is well-known that steady states are signaled by terminals in asynchronous boolean networks. Then, it becomes worth to use these models to proceed with a prelim-

inary study. We point out that in a biological context, the concept *attractor* is more often used than "terminal". Therefore, we use the term "attractor" instead of "terminal" when we refer to this concept in biological boolean networks.

It is not difficult to see that the number of states of these models grows exponentially with the number of components of the system. For example, a model which considers 10 components admits 2^{10} states. Because of this, and since the most of the biological models admits much different components (usually, much more than 10), it becomes important to both develop tools to minimize these boolean network models and new ideas to find features like attractor with few computational cost. In order to do this, we propose to take into account the ideas already used in automata theory. Although we do not present any new algorithm, this work paves the way to new approaches to this problem.

In this paper, we apply the concept of bisimulation to propose a new method to preliminarily study these biological systems. Bisimulations are already used in several minimization processes. Furthermore, the possibility of combining bisimulation with modal logic which admit modalities (see [1]) turns out that it is a powerful tool to study state transition systems. The usage of modal logic is possible due to the possibility of interpreting biological boolean network as Kripke models. However, in this paper, we will only propose bisimulations to develop new minimization processes which allow us to find the attractors of boolean networks and we do not consider any background logic. Thus, given a digraph (V, E), we say that $\mathcal{S} \subseteq V \times V$ is a bisimulation if $\mathcal{S}$ is not empty, and if it is an equivalence relation such that:

– If $(v, w) \in \mathcal{S}$ and $(v, v') \in E$ then there exists $w' \in V$ such that $(w, w') \in E$ and $(v', w') \in \mathcal{S}$.

– If $(v, w) \in \mathcal{S}$ and $(w, w') \in E$ then there exists $v' \in V$ such that $(v, v') \in E$ and $(v', w') \in \mathcal{S}$.

Outline. We begin by presenting some definitions and a theorem that relates attractors with bisimulations. Then, we enhance the difference between minimizing boolean networks using bisimulation and other known methodologies used in the study of such systems. Finally, we present some conclusions and directions to follow.

2 Bisimulations and Attractors

The dynamics within a cell are guided by several components: proteins, RNA, genes, ribosomes, *etc...* Each of these components induces or inhibits the production/activation of some of the other ones. Thus, it is very difficult to understand a biological regulatory network and, usually, only some main features are studied. As referred, one of these features are the steady states. Steady states are related to the modes of operating of a cell. For instance, in [4], a model for *E. coli* with two steady states is presented: one is related to a configuration in which the organism metabolizes sugar, grows and replicate itself; and the other relates to a configuration in which the organism does not metabolize sugar, does not grow and does not replicate itself.

There exists results which shows that the study of asynchronous boolean models can be used to identify steady states (see [9]). In practice, each terminal represents a steady state. We follow with a formal definition of terminal.

Definition 1. Let (V, E) be a graph.

We say that $v \in V$ has a transition to $w \in V$ and we write $v \to w$, if $(v, w) \in E$. We write $v \nrightarrow w$ otherwise.

We say that there is a *path* from v to w if there exist v_1, ..., $v_n \in V$ such that $v \to v_1$, $v_1 \to v_2$, ..., $v_n \to w$.

A Strongly Connected Component (SCC) is a subset A of V such that there is a path between any two element of A.

A set A is a terminal if it is a SCC and $\nexists\ a \in A, v \in V \backslash A$: $a \to v$.

We point out that, since the biological boolean models represent a finite number of components of a cell, in this section, we assume that all the considered digraphs are finite, *i.e.*, for any digraph (V, E), $|V| < \infty$.

To find the attractors of digraphs, several methodologies can be used. In [7], [10] some methods are presented. However, here, we present some new ideas that can lead to a new approach on this theme, based in bisimulations. We thus follow with a definition of a particular class of bisimulations and we present and prove a theorem that relates this class of bisimulations with attractors.

Definition 2. Let (V, E) be a graph.

We say that $\mathcal{B} \subseteq V \times V$ is a complete bisimulation if it is a bisimulation and there exists $B \subseteq V$ such that $\mathcal{B} = B \times B$ (any two elements of B are related).

We say that a complete bisimulation $\mathcal{B}$ is minimal if there is not any other complete bisimulation $\mathcal{B}'$ such that $\mathcal{B}' \subsetneq \mathcal{B}$.

Lemma 1. *Let (V, E) be a graph, $B \subseteq V$ and $\mathcal{B} = B \times B$ a minimal complete bisimulation. For any $A \subsetneq B$, $\exists\ a \in A, v \in B \backslash A$ such that $a \to v$.*

Proof

Let us assume that there exists $A \subsetneq B$ such that, for any $a \in A, v \in B \backslash A$, $a \nrightarrow v$.

In this case, we can easily verify that $A \times A$ is an equivalence relation since all states of A are related. By hypothesis, for any $(a, a') \in A \times A \subseteq \mathcal{B}$ such that $a \to b$, there exists some b' which verifies $a' \to b'$ and $(b, b') \in \mathcal{B}$. Since for any $a \in A, v \in B \backslash A$, $a \nrightarrow v$, we can conclude that $b, b' \in A$ and, therefore, $(b, b') \in A \times A$. Thus, $A \times A$ is a complete bisimulation and this contradicts the minimality of $\mathcal{B}$.

□

Theorem 1 *Let (V, E) be a graph. $\mathcal{B} = B \times B \subseteq V \times V$ is a minimal complete bisimulation $\Leftrightarrow$ B is a terminal of V.*

Proof

" $\Rightarrow$ "

We start by proving that if $\mathcal{B}$ is a minimal complete bisimulation, then there exists a path between any two elements of B. We prove that B is a terminal afterwards.

We consider $u, v \in B$. By Lemma 1., we know that there is a transition $u \to u_1$ from $\{u\}$ to $B\backslash\{u\}$. If $u_1 = v$, we are done. Otherwise, using Lemma 1. again, we know that there is a transition from $\{u, u_1\}$ to some $u_2 \in B\backslash\{u, u_1\}$. Here, either $u \to u_2$ or $u_1 \to u_2$. In any case, there is a path from u to u_2. Again, if u_2=v we are done. Otherwise, we can continue to apply this procedure till find a path between u and v (this procedure will end in finite time since we are only considering finite graphs). As u and v were arbitrary, we can conclude that B is a SCC.

Finally, if $u \in B$ and $u \to v$, then $(u, u) \in \mathcal{B}$ and, by definition of complete bisimulation, $(v, v) \in \mathcal{B}$. Then $v \in B$ and, thus, B is a terminal.

" $\Leftarrow$ "

We now assume that B is a terminal of V. We can easily see that $\mathcal{B} = B \times B$ is a equivalence relation since all states are related. We consider $(u, v) \in \mathcal{B}$ and $u \to u'$. Since B is a terminal, $u' \in B$ and $\exists v' \in B$ such that $v \to v'$. Furthermore, $(v, v') \in \mathcal{B}$, by definition and, therefore, $\mathcal{B}$ is a complete bisimulation.

Let us assume that $\mathcal{B}$ is not minimal, *i.e.*, there is a complete bisimulation $\mathcal{A} := A \times A \subsetneq \mathcal{B}$. Since B is terminal, it is possible to find a path from any $a \in A$ for any $b \in B\backslash A$. Thus, $\exists a' \in A, b' \in B\backslash A$ such that $a' \to b'$. But this contradicts the fact of $\mathcal{A}$ being bisimulation because $b' \notin A$.

□

This theorem can help us to develop a new minimization methodology which preserves the attractors without even know them. The general idea is to find a bisimulation $\mathcal{B}$ such that any complete bisimulation contained in $\mathcal{B}$ is minimal. Thus, we can compute the "quotient digraph" in order to obtain a minimized model in which the states of the attractors may be "clustered". Nevertheless, these attractors are individually preserved. We follow with two examples in order better understand how the minimization *via* bisimulation is made.

In this example, we consider two asynchronous boolean networks models. In this kind of boolean network, the directed edges representing the transitions between states are defined according to some boolean equations. However, we can only update the value of one variable at each time. To simplify, we do not consider any loops.

For the first example, we pick a purely theoretical model which is presented in Figure 1. This example is presented in order to further distinguish two methodologies. In this figure, the attractor of the model is enhanced by an orange box. Although this model is theoretical, it could result from a system comporting three components (a, b and c) whose state transition is computed by updating the value of a single component at each time and according to the following boolean equations:

$$\begin{cases} a := a \vee b \\ b := a \vee (b \oplus c) \\ c := (a \oplus \neg b) \vee c \end{cases}$$

In these equations, $\oplus$ is the XOR boolean operator.

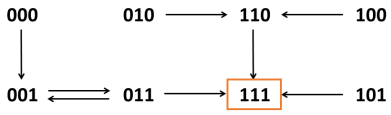

Fig. 1. Model with an attractor composed of a single state.

In order to minimize this model, we can find a bisimulation $\mathcal{B} = \{(000, 000), (001, 001), (011, 011), (111, 111), (010, 010), (110, 110), (100, 100), (101, 101), (001, 011), (011, 001), (110, 101), (101, 110), (010, 100), (100, 010)\}$. It is not difficult to verify that all complete bisimulation contained in this bisimulation are minimal. Thus, we can construct the "quotient digraph" by clustering the states in the same equivalence class. The transitions of the "quotient digraph" are introduced by the following rule: "If $a \to b$, then, $[a] \to [b]$ (where $[a]$ and $[b]$ are the equivalence classes of a and b, respectively)". This quotient digraph is then presented in Figure 2.

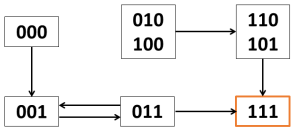

Fig. 2. Quotient digraph of the model in Figure 1.

We now consider a real example. In [3], it is presented a biological system that regulates the circadian rhythm in a cyanobacteria, *i.e.* this system models the biological processes (which are periodic and whose period is 24 hours) that regulates the perception of a day-cycle by an organism. In [3], this system is studied with a asynchronous boolean network and the attractors of the resulting network are found and it is studied the robustness of this model. The asynchronous boolean network used is defined by the boolean equations which follow:

$$\begin{cases} a := \neg s \\ s := ts \\ t := a \\ ts := t \wedge a \end{cases}$$

This model describes the dynamics of the three phosphorylated forms of Kai C (using boolean variables t, ts, and s) and the protein Kai A (with a boolean variable a). These four form are responsible for the core of the cyanobacterial circadian clock. The referred system is modeled by the presented boolean functions and the resulting asynchronous boolean network, whose attractor is enhanced by an orange box, is shown in Figure 3.

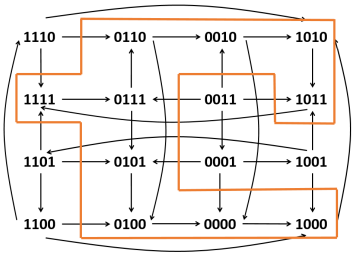

Fig. 3. Model of the circadian rhythm in a cyanobacteria.

As before, we can find a bisimulation such that all complete bisimulations contained on it are minimal and compute the "quotient digraph". Hence, if we consider the following bisimulation $\mathcal{B} = \{(a, b) : a, b \in \{0110, 0010, 1010, 1111, 0111, 1011, 0101, 0100, 0000, 1000\}\} \cup \{(a, a) : a \in \{1110, 1101, 1100, 0011, 0001, 1001\}\}$, the resultant quotient digraph is the one which is presented in Figure 4.

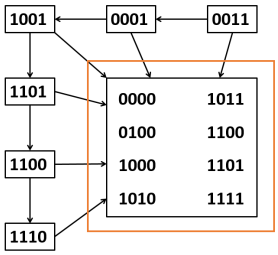

Fig. 4. Minimized model for the circadian rhythm in a cyanobacteria.

In both cases, we can see that we obtain a minimized model in which the attractors are preserved. This is due to the fact that all complete bisimulations contained in the bisimulation used are minimal.

3 Comparing Bisimulation with other Reducing Methods

In this section, we compare this method of minimization with other methodologies which are commonly used in this field. We do this in order to distinguish our minimizing method from those and we point out some advantages (and disadvantages) of our method when comparing to other approaches.

Firstly, we compare our quotient digraph with the hierarchical representations. This kind of representations sort the states of a model according to their distance to the attractors. This approach is widely used in the study of systems. However, the main disadvantage is that one must know *a priori* which are the attractors of the model in order to obtain a hierarchical representation. It is not difficult to see that our method can only get together two states whose distance to the attractors is the same (since transitions are someway preserved). However it only clusters states whose behavior is similar. To illustrate this we call the minimized model in Figure 4. In this example, if we consider any state which is not part of the attractor, we can see that its distance to the attractor is "1". Despite this fact, we can see that any two of these states are not clustered because each one of them presents a different behavior. Actually it may be important to study these differences in their behavior. Actually, they allow us to discover the longest possible "path" to the attractor.

Another widely used method to minimize boolean networks and, then, search for attractors is clustering the SCCs. This allows us to minimize a model and still preserve the attractors. This is a well-known idea and several other methods to find attractor where developed after it (for instance, see [10]). On one hand, the method which we present can clusters sets of states which are not in the same SCC and, therefore, it clusters states which would not be clustered when we cluster the SCC's. For instance, recalling the example in Figure 2, we can see that the states 010 and 100 were clustered and were not in the same SCC; on the other hand, constructing the quotient digraph, it can happen that we do not cluster all SCC's. For example, the quotient digraph presented in Figure 2 still has a SCC. This is due to the fact that our method can only cluster SCCs which are terminals. We can see this because, for any SCC which is not a terminal, there exists some state in the SCC that admits a transition for a state out of that SCC. Therefore, it may be impossible to find a complete bisimulation to cluster all its states.

Finally, we point out a last important feature of bisimulations. Since we are dealing with discrete state transition models (automata), it can be useful to use modal logic to reason about such models. Hence, it could be useful to obtain minimization processes which guarantee that all states in a cluster verify the same modal formulas. Indeed, due to their definition, bisimulations are suitable to be used with modal logic. More information about this can be found in [1].

4 Conclusions

Bisimulations can be used to obtain to minimize biological boolean models and, guaranteeing some conditions, the methodology we presented preserves the attractors. Although we present only the main ideas and some examples of the application of this method, it can provide the basis for a new minimization algorithm. Actually, in future, we are planing to develop this complete algorithm which applies these ideas to minimize biological boolean models.

We also evaluated the convenience of using this minimization methodology when compared with other methods already used. It provides a new way of looking at biological models and it can be useful in their study. When comparing with other methods, it has both some advantages and disadvantages. However, as seen, since it preserves the attractors and, moreover, it can be combined with a modal logic, we believe that this approach is worth.

In future, we also plan to study how can modal logic fit in these biological boolean models and, if possible, to find an axiomatization of such systems which would allow us to formally prove diverse properties of them. Actually, in continuous models which use ODEs, it was applied a dynamic logic (which integrates first-order features) proposed by A. Platzer – Differential Dynamic Logic (see [8]) – to formally reason about them. Some initial work can be found in [6]. We believe that is possible to obtain a similar results in discrete models and, in particular, in boolean networks.

References

1. Blackburn, P., De Rijke, M., Venema, Y.: Modal Logic: Graph. Darst, vol. 53. Cambridge University Press (2002)
2. Chaves, M.: Predictive analysis of dynamical systems: combining discrete and continuous formalisms. Ph.D. thesis, Gipsa-lab (2013)
3. Chaves, M., Preto, M.: Hierarchy of models: From qualitative to quantitative analysis of circadian rhythms in cyanobacteria. Chaos: An Interdisciplinary Journal of Nonlinear Science 23(2), 025113 (2013)
4. Chaves, M., Tournier, L.: Predicting the asymptotic dynamics of large biological networks by interconnections of boolean modules. In: Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on. pp. 3026–3031. IEEE (2011)
5. De Jong, H.: Modeling and simulation of genetic regulatory systems: a literature review. Journal of computational biology 9(1), 67–103 (2002)
6. Figueiredo, D.: Differential dynamic logic and applications. Master's thesis, University of Aveiro (2015)
7. Naldi, A., Remy, E., Thieffry, D., Chaouiya, C.: A reduction of logical regulatory graphs preserving essential dynamical properties. In: Computational Methods in Systems Biology. pp. 266–280. Springer (2009)
8. Platzer, A.: Logical analysis of hybrid systems: proving theorems for complex dynamics. Springer Science & Business Media (2010)
9. Thomas, R., Kaufman, M.: Multistationarity, the basis of cell differentiation and memory. ii. logical analysis of regulatory networks in terms of feedback circuits. Chaos: An Interdisciplinary Journal of Nonlinear Science 11(1), 180–195 (2001)

10. Tournier, L., Chaves, M.: Interconnection of asynchronous boolean networks, asymptotic and transient dynamics. Automatica 49(4), 884–893 (2013)